



Asking New Questions with Old Data: The Centralized Open-Access Rehabilitation Database for Stroke

Keith R. Lohse^{1*}, Sydney Y. Schaefer^{2,3}, Adam C. Raikes³, Lara A. Boyd⁴ and Catherine E. Lang^{5,6,7}

¹ School of Kinesiology, Auburn University, Auburn, AL, USA, ² School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA, ³ Department of Health, Physical Education and Recreation, Utah State University, Logan, UT, USA, ⁴ Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada, ⁵ Program in Physical Therapy, Washington University School of Medicine in St. Louis, St. Louis, MO, USA, ⁶ Program in Occupational Therapy, Washington University School of Medicine in St. Louis, St. Louis, MO, USA, ⁷ Department of Neurology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA

Background: This paper introduces a tool for streamlining data integration in rehabilitation science, the Centralized Open-Access Rehabilitation database for Stroke (SCOAR), which allows researchers to quickly visualize relationships among variables, efficiently share data, generate hypotheses, and enhance clinical trial design.

Methods: Bibliographic databases were searched according to inclusion criteria leaving 2,892 titles that were further screened to 514 manuscripts to be screened by full text, leaving 215 randomized controlled trials (RCTs) in the database (489 independent groups representing 12,847 patients). Demographic, methodological, and statistical data were extracted by independent coders and entered into SCOAR.

Results: Trial data came from 114 locations in 27 different countries and represented patients with a wide range of ages, 62 year [41; 85] [shown as median (range)] and at various stages of recovery following their stroke, 141 days [1; 3372]. There was considerable variation in the dose of therapy that patients received, 20 h [0; 221], over interventions of different durations, 28 days [10; 365]. There was also a lack of common data elements (CDEs) across trials, but this lack of CDEs was most pronounced for baseline assessments of patient impairment and severity of stroke.

Conclusion: Data integration across hundreds of RCTs allows clinicians and researchers to quickly visualize data from the history of the field and lays the foundation for making SCOAR a living database to which researchers can upload new data as trial results are published. SCOAR is a useful tool for clinicians and researchers that will facilitate data visualization, data sharing, the finding of relevant past studies, and the design of clinical trials by enabling more accurate and comprehensive power analyses. Furthermore, these data speak to the need for CDEs specific to stroke rehabilitation in randomized controlled trials.

PROSPERO 2014:CRD42014009010

Keywords: stroke, rehabilitation, informatics

OPEN ACCESS

Edited by:

Thanh G. Phan,
Monash Health, Australia

Reviewed by:

Phil Clatworthy,
University of Bristol, UK
Muhib Khan,
Michigan State University, USA

*Correspondence:

Keith R. Lohse
rehabinformatics@gmail.com

Specialty section:

This article was submitted to Stroke,
a section of the journal
Frontiers in Neurology

Received: 23 April 2016

Accepted: 05 September 2016

Published: 20 September 2016

Citation:

Lohse KR, Schaefer SY, Raikes AC,
Boyd LA and Lang CE (2016) Asking
New Questions with Old Data: The
Centralized Open-Access
Rehabilitation Database for Stroke.
Front. Neurol. 7:153.
doi: 10.3389/fneur.2016.00153

INTRODUCTION

The information architecture in rehabilitation science is poor (1). For example, randomized controlled trials (RCTs) are the basic “unit” of information that guide clinical practice. Yet when clinicians and scientists want to ask a very basic question of these data, they are published: (1) across a wide spectrum of journals and formats that often have limited access (e.g., payment required for access); (2) embedded potentially in text, tables, figures, or even supplemental materials; and (3) with very few common data elements (CDEs) reported across studies (2, 3). Thus, despite the tremendous time and financial burdens associated with even a single RCT, the resultant data lack a consistent structure. This lack of structure is an unnecessary barrier to integration in future scientific and clinical practice. Efforts to streamline data integration should increase the transparency and visibility of comprehensive bodies of evidence, rather than a single study, to better inform clinically relevant questions such as, “How do therapy outcomes change with increased time in therapy?” or “How variable are outcomes, historically, for specific parameters of therapy?”

We now introduce one such tool for streamlining data integration: the Centralized Open-Access Rehabilitation database for Stroke (SCOAR). In short, SCOAR is a central repository for summary statistics from RCTs. SCOAR currently contains data from a systematic review and extraction of papers from 1981 to early 2014 (described in detail below), but the goal of SCOAR is much bigger: to create a “living” database where new data can be added as clinical trials are completed. Imposing such an architecture (4) on clinical trial data would allow basic and clinical scientists to (1) quickly and easily visualize relationships among variables, (2) efficiently share data, (3) generate hypotheses based on noticeable patterns or even “gaps” in the current data, (4) search the current literature from the data up (rather than key-terms down), and (5) improve clinical trial design through more accurate and comprehensive power analyses.

Generally speaking, the goal of SCOAR is to improve the design of future clinical trials by giving researchers fast and easy access to the historical range of effect-sizes, based on thousands of stroke patients who received therapies of different types, different doses, at different times, and were measured on different outcomes. From our perspective, the effort associated with the design, implementation, and dissemination of randomized clinical trials *deserves* an information architecture that supports and increases their visibility. In the current paper, we (1) explain the systematic search and data extraction that led to the creation of SCOAR; (2) present summary statistics for the major variables in SCOAR, including the geographical reach, to understand how SCOAR data represent research in stroke rehabilitation; and (3) based on the lack of CDEs we find across many variables, we argue for a consistent set of CDEs in rehabilitation trials (CDEs to describe participants, methodology, and outcomes). SCOAR lays the foundation for an information architecture that captures some of the complex and multivariate nature of neurorehabilitation. Most importantly, this information architecture is scalable, making it easy to add new data as new trials are published.

MATERIALS AND METHODS

Systematic Review

A systematic search was undertaken in May 2014 (full details in PROSPERO 2014:CRD42014009010) using the following databases: MEDLINE, EMBASE, Cochrane CENTRAL, Cochrane-CDSR, and CINAHL. Outside references were also incorporated from previous reviews (5, 6) (**Figure 1**). After removing duplicates, 2,892 remaining titles and abstracts were screened by independent coders (two pairs of trained graduate students) based on the following inclusion/exclusion criteria (7):

Population

Human adults with stroke, >18 years of age, with a motor impairment as a result of stroke. Any etiology was included (e.g., ischemic, hemorrhagic, sub-arachnoid hemorrhage) provided that the study identified patients as having a stroke or cerebrovascular accident. Groups with mixed neurological impairments (e.g., patients had either stroke or TBI) were excluded. (In the RCTs we found, the majority of groups were characterized as having first-ever unilateral strokes, without other neurological conditions, but these were not criteria for inclusion.)

Intervention

Any physical or occupational therapy interventions that required active movement on the part of the participant were included. RCTs that used stimulation techniques were allowed (e.g., FES, TMS, tDCS) if combined with active movement. Studies with strictly pharmaceutical interventions (e.g., to treat spasticity) were excluded.

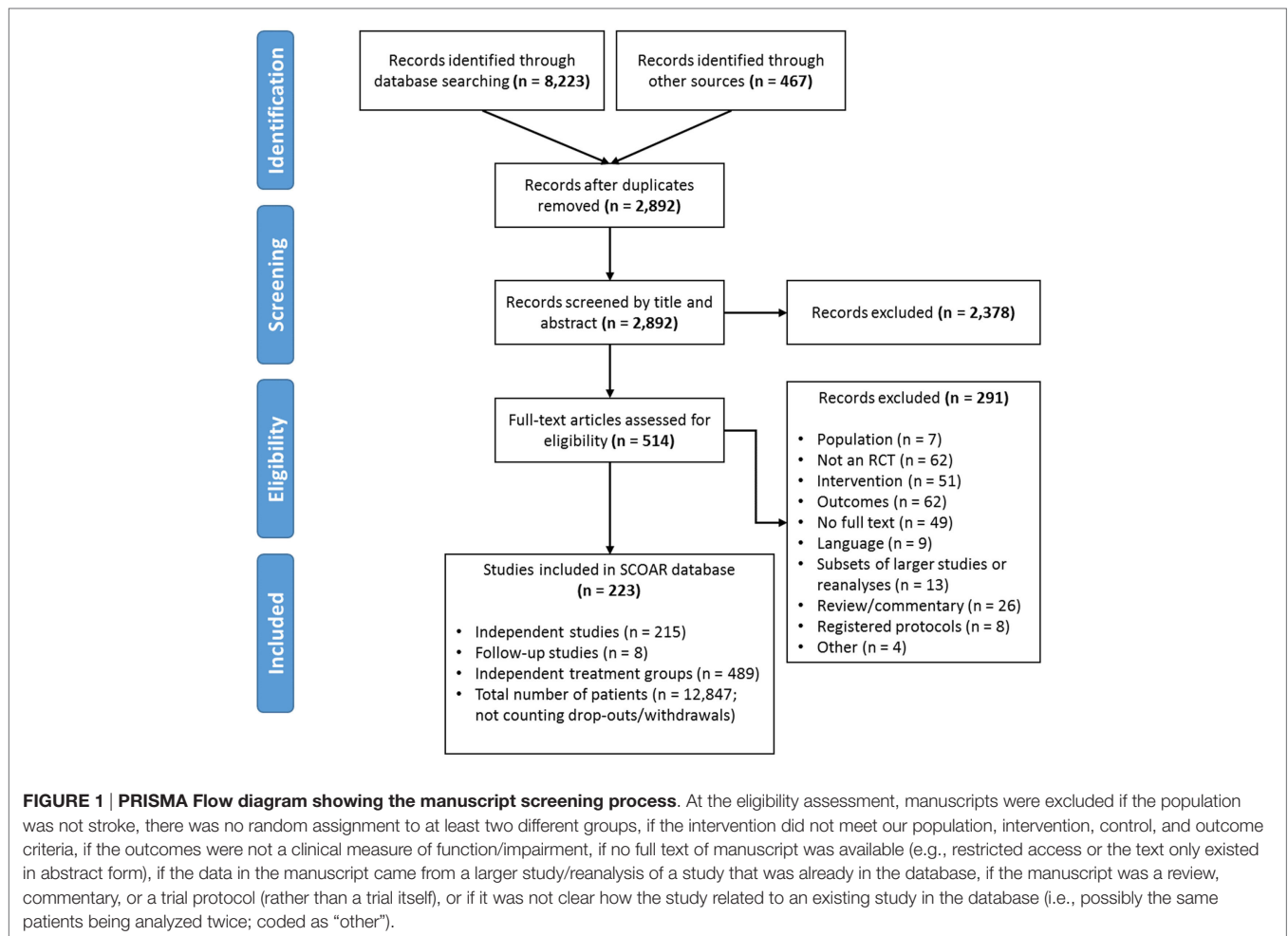
Control

All studies had to be RCTs and studies were required to explicitly state random assignment to groups. The condition nominally identified as “control” by the authors was coded as control, or if a group received “conventional care,” “routine therapy,” or “standard care” without being specifically named as control, it was assumed that this was the control condition. All other conditions were labeled as “experimental.”

Outcome

Only empirically validated assessments of impairment or functional motor capacity (i.e., activity limitations) administered by the clinician were taken as outcomes (e.g., no self-report measures, no neuro-imaging measures, no study-specific kinematic/kinetic measures). Note that studies could include these other types of measures, but only measures of impairment or function were extracted for inclusion in the database at this time.

If there was disagreement between the coders, the first author (Keith R. Lohse) provided a tie-breaking vote on inclusion or led discussion until agreement was reached. The same PICO inclusion/exclusion criteria were also applied to the full text review. During the full text review, the authors systematically extracted data from the manuscripts (described below in Section “Data Extraction”). During the extraction, concerns emerged regarding how to consider statistically related data within SCOAR. In other words, SCOAR is designed to



longitudinally track groups over time (viz., from baseline assessment, to terminal assessment, to the last long-term follow-up) and, therefore, studies with long-term follow-up data, subsets of data, or re-analyzed data published in separate manuscripts presented a unique challenge. To maintain independence of the data for statistical analysis, any given group of patients should only appear in the SCOAR database once. Thus, subsets and re-analyses were excluded ($n = 8$) or if the relationship between manuscripts was unclear ($n = 5$), those studies were excluded as well. However, if a paper published a long-term follow-up or reported usable outcome measures in a separate manuscript ($n = 8$), then these data were grouped together in the SCOAR database [e.g., Ref. (8–11)], such that independence between groups is maintained. As shown in **Figure 1**, there were a number of studies for which the full text could not be obtained ($n = 49$), or even if the full text was available, there was no available English translation ($n = 9$). (At present, no effort has been made to translate these non-English texts, but with adequate translation, the data from these studies may be eligible for inclusion following review.) In total, SCOAR has summary statistics from 12,847 patients in 489 independent groups (as of 2016-03-31).

Data Extraction

Separate from the pairs of graduate students who screened by title and abstract, three authors (Keith R. Lohse, Sydney Y. Schaefer, Adam C. Raikes) extracted data by hand from electronic copies of manuscripts using a standardized data extraction form. These extracted values were then entered by hand into a spreadsheet. Discrepancies between extractors were discussed until consensus was reached. A copy of the data extraction form is listed in Data Sheet S1 in Supplementary Material and an explanation of all current SCOAR variables is included in the data dictionary, Data Sheet S2 in Supplementary Material.

One author (Keith R. Lohse) extracted outcome data for either the primary outcome or the first usable outcome, when no primary outcome was stated or the primary outcome was not usable. As per the PICO criteria, a usable outcome measure was defined as an assessment given by a therapist that was a clinical measure of motor impairment and/or function. Two authors (Sydney Y. Schaefer, Adam C. Raikes) also reviewed all of the articles to extract all additional data related to the Fugl-Meyer Assessment (FMA, upper-extremity portion, a measure of upper limb impairment) and gait speed (including the 10-m walk test, or variations thereof, a measure of walking function). This second

set of outcomes was extracted regardless of whether the FMA or gait speed measure was stated as a primary outcome. FMA and gait speed measures were chosen for this second extraction because they were the most commonly reported measures for those two domains in previous work (12).

In extracting these data, the relevant outcome data were often clearly presented. For instance, baseline means and SDs were consistently reported in the text (i.e., the pre-intervention assessment, reported in 93 and 83% of cases), whereas terminal means (i.e., the most immediate post-intervention assessment; calculable for 93% of cases) sometimes had to be calculated by the data extractors from reported change scores (12% of calculable cases). Additionally, terminal SDs (calculable for 72% of cases) were sometimes estimated by the data extractors from figures or reported confidence intervals (3% of calculable cases).

Other data were often not reported in the text and, therefore, had to be estimated by the data extractors. For instance, the time scheduled for therapy was typically estimated from written descriptions (e.g., “1 h per day, 5 days per week for 4 weeks” = 20 h of scheduled therapy). An estimate of the time scheduled for therapy was calculable in 74% of cases. Although some studies reported repetitions of movements (13, 14) or active time in therapy (15, 16), the most common metric was the time scheduled for therapy. Although time scheduled for therapy is likely to be a poor indicator of the actual amount of therapy received (17), it is positively correlated with the amount of therapy received and was the most common metric reported across trials. Constraint-induced movement therapies (18, 19) or other “forced-use” therapies (20) were problematic for estimating the time scheduled for therapy because they often do not specify the amount of time actually spent in constraint or forced-use. In line with previous work (6), we calculated time under constraint in three different ways: (1) counting 100% of constraint time as time in therapy, (2) counting 50% of constraint time as time in therapy, and (3) counting 0% of constraint time as time in therapy. The 50% time calculation is preferred because it has the most plausible assumptions (i.e., some, but not all, of constraint time is spent using the affected extremity), but all three calculations are available to researchers in the database. We further note that some constraint studies reported constraint for a “percentage of waking hours,” and these percentages were converted to hours based on 16 waking hours per day.

Estimation of Within-Group Effect Size

Compared to effect-size calculations in other rehabilitation meta-analyses (5, 6), the default effect-size in SCOAR represents the change within groups over time rather than a difference between groups at a specific time point. Calculating a standardized effect-size for within-group change is important, because these effect-sizes allow for the greatest flexibility in integrating changes across studies using the most data [i.e., each group’s improvement (or decrement) is normalized to their baseline at the beginning of the intervention]. Calculation of these within-group changes creates some unique challenges for meta-analysis (see below), but

all effect-size calculations were for a Cohen’s d as described in Borenstein et al. (21).

$$d = \frac{\bar{y}_1 - \bar{y}_2}{s_{\text{pooled}}} \quad (1)$$

where s_{pooled} is the between-person SD pooled between the two different time points to create a single estimate of the between-person variance.

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}} \quad (2)$$

Thus, s_1^2 and s_2^2 refer to the variance at the baseline and the terminal assessment, respectively. In 28% of the total cases, the variance was not estimable at the terminal assessment. However, in 16% of the total cases the baseline SD was available and used in the calculation of d when the terminal SD was not available. In fewer cases (1% of total cases), the variance at baseline was zero (e.g., all participants had a Functional Ambulation Category of zero due to a floor effect in the outcome measure). In those cases, the terminal SD was used in the calculation of d .

Finally, given the wide range of the effect-sizes and sample sizes we observed in the data, we transformed these effect-sizes from Cohen’s d to Hedges’ g . Cohen’s d is biased to overestimate the underlying effect-size in small samples whereas Hedges’ g is a more conservative and unbiased calculation of the effect-size in which the d value is reduced proportional to the sample size (21). Subtraction in the effect-size calculations was arranged such that positive values in SCOAR always reflect improvement relative to baseline.

Although the default effect-sizes in SCOAR represent the normalized improvement within a group over time, we should point out that SCOAR also contains the sample size, mean, and SD for all groups at the baseline, terminal, and follow-up assessment (if applicable). Having these descriptive statistics for each group at each time point allows researchers to readily calculate between-group effect sizes if those effects are more relevant to their research question. Thus, by extracting the sample size, mean, and SD at each time point, SCOAR allows researchers to calculate outcomes in three different ways: (1) a standardized effect-size showing change within a group over time (the default SCOAR effect-size), (2) a standardized effect-size showing the difference between groups at a single point in time, or (3) the original “raw” units of the outcome measure. Although using original units precludes combining different outcomes into a single analysis, this is a sensible option when restricting outcomes to a single type (i.e., measures of gait speed can all be expressed in terms of meters per second, so there is no need to normalize).

Estimation of the Correlation between Time Points and Effect-Size Variance

In order to conduct quantitative meta-analyses with these data, we also need to calculate the variance of the individual effect-sizes. For statistically dependent “within-subject” data, the correlation

between time points, r , is required for an accurate estimation of effect-size variance, V_d (21).

$$V_d = \left(\frac{1}{n} + \frac{d^2}{2n} \right) 2(1-r) \quad (3)$$

The correlation between baseline and terminal (or follow-up) scores was never reported in any of the included RCTs. We were, however, able to estimate the correlation between baseline and terminal assessments from studies that provided either (A) individual patient data or (B) SDs of the baseline, terminal, and baseline-to-terminal change scores (21). In this subset of studies (13, 22–34), we observed that the median correlation was $r = 0.87$, IQR = (0.70, 0.93) and the minimum correlation was $r = 0.28$ between baseline and terminal assessments. As such, we calculated effect-size variance based on the conservative assumption that $r = 0.5$ for all studies (which generally creates larger estimates of variance in the data, widening our confidence intervals). This estimated correlation is, however, an easy value for researchers to manipulate within SCOAR and recalculate their own effect-size variances. Thus, the default effect-size variance in SCOAR is calculated based on $r = 0.5$, which we consider to be a conservative estimate, but researchers can easily recalculate effect-size variances by scaling this correlation up or down.

RESULTS

Descriptive statistics across the 489 independent groups of participants currently in the database are listed in **Table 1**. The full database (as of 2016-03-31), the data dictionary, a full reference list all trials, and the Creative Commons license for SCOAR are

available from <https://github.com/keithlohse/SCOAR>. (Note that SCOAR is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License by Keith R. Lohse.)

As shown in **Table 1**, the data in SCOAR are representative of a wide range of patients (in terms of day-post stroke and age) and different types of interventions (in terms of sample size, duration of the intervention, and dose of therapy given over the intervention). An important point to note are the different methods used for tracking the dose of therapy patients received. Consistent with our previous work (6), the most common metric for tracking dose was the hours of scheduled therapy. The best measures of dose are the time actively spent doing therapy (19, 35) (which was more commonly reported in gait therapies) or the actual repetitions of therapy exercises (14, 36) (which was more commonly reported in trials using robotic assistive devices).

Geographical Distribution of Studies

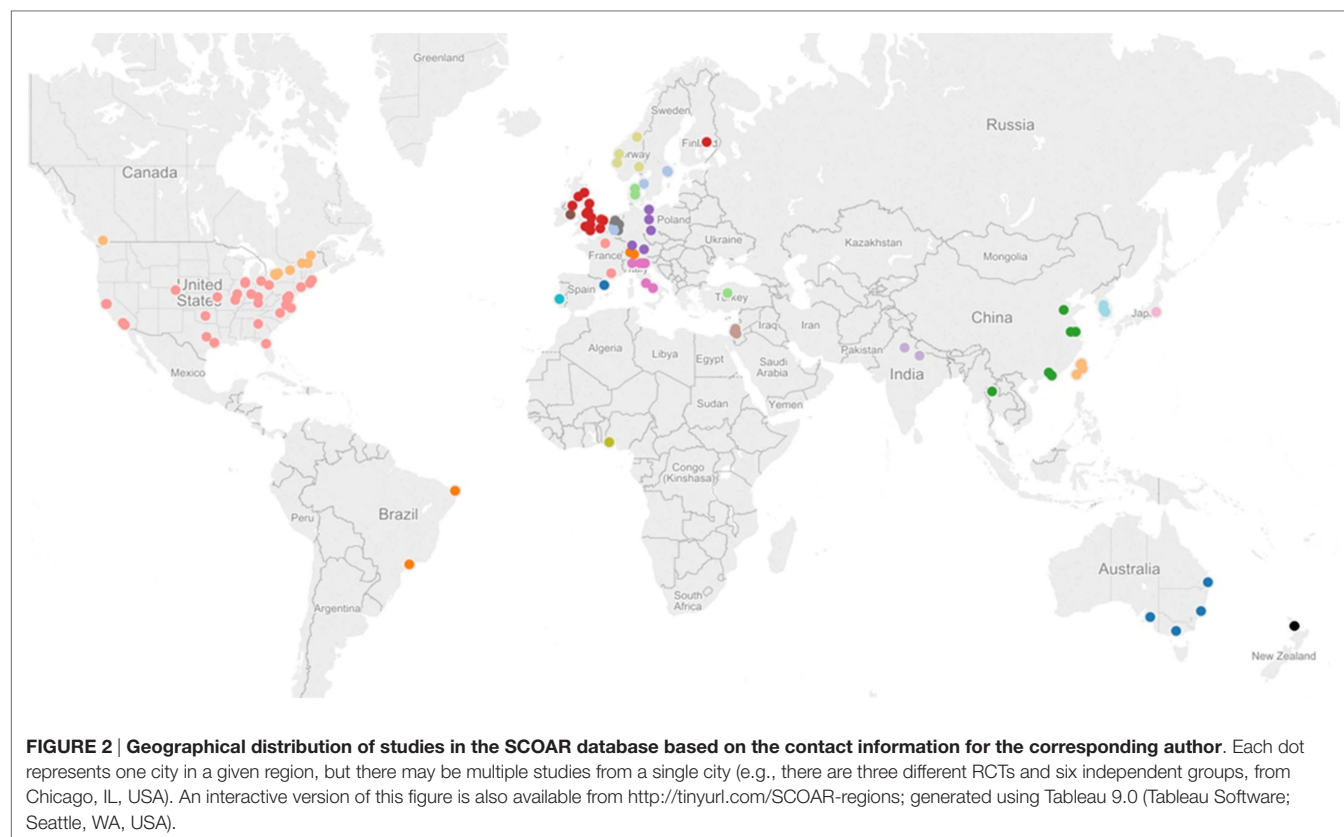
As shown in **Figure 2**, trials in the SCOAR database come from a wide range of countries. The location of a trial was estimated based on the contact information for the corresponding author. Note that multiple studies may be represented by each dot (if trials were conducted at the same location), so we refer readers to the interactive version of this figure¹ where they can see the number of trials, number of groups, and the references for each location. Overall, trials came from 114 locations in 27 different countries. From this wide range of trial locations, it does not appear that SCOAR is geographically biased relative to the larger population of stroke rehabilitation trials. The extent to which the population of trials might be biased, however, is an important

¹<http://tinyurl.com/SCOAR-regions>

TABLE 1 | Demographic statistics for the studies included in SCOAR.

Variable	Mean (SD)	Median (IQR)	Min; Max	Groups with missing values
Mean patient age (years)	62.6 (6.7)	62.4 (57.3; 67.3)	41.30; 85.20	12
Mean days post stroke (days)	509 (652)	141 (31; 840)	1; 3,372	22
Duration of intervention (days)	45 (39)	28 (28; 42)	10; 365	20
Estimated Time Scheduled for Therapy (h)				
Max time calculation	34.2 (45.3)	20.0 (10.0; 36.1)	0.0; 280.0	129
50% time calculation	29.9 (31.6)	20.0 (10.0; 36.0)	0.0; 220.7	129
Min time calculation	25.5 (24.5)	20.0 (10.0; 32.0)	0.0; 220.7	129
Method for Tracking Dose of Therapy – N (% out of 489)				
Hours scheduled	307 (63%)			
Time in therapy	26 (5%)			
Active time	37 (8%)			
Repetitions	20 (4%)			
Groups Reporting ITT Analysis	158 (32%)			
N per Group at baseline calculation	26 (26)	18 (11; 31)	4; 165	0
N per group at terminal calculation	25 (25)	17 (10; 30)	4; 165	6
N per Group at follow-up calculation	29 (29)	19 (2; 34)	4; 165	215
Time from baseline to follow-up (days)	178 (137)	180 (88; 206)	31; 1,098	231
Coded as experimental groups	285 (58%)			
Coded as control groups	204 (42%)			

N per group refers to the number of participants whose data were used in the calculation of the mean and the SD at each time point, not necessarily the number of participants enrolled in/actively participating in the study at that time point (depending on the analytical method used by the authors). Time in days and N per group are rounded to the nearest integer. ITT = intention to treat.



question. As seen in **Figure 2**, there is a dearth of trials for regions such as South America, Africa, central Asia, and eastern Europe.

Common Data Elements for Baseline Assessment and Initial Severity

Across the 489 independent groups of participants, there was a wide range of data elements, but unfortunately there were relatively few CDEs reported across the different studies. This lack of CDEs was especially pronounced for variables measuring the baseline severity of stroke or patient impairment. For example, data describing cognitive status with the Mini-Mental Status Exam (MMSE) were reported for only 152 groups in some form (31% of cases). A mean or median MMSE score was given for 82 (17%) of these groups, whereas 70 (14%) of these groups only reported some cut-off/critical MMSE value in their inclusion/exclusion criteria. The MMSE example was not unusual, however, and was actually the most common CDE for baseline assessment. For a list of the most common baseline assessments and their prevalence (see **Table 2**).

This lack of CDEs (or at least under-reporting of CDEs) is undoubtedly a major barrier to data integration and meta-analysis, especially for ostensibly common baseline assessments, such as the FIM or NIHSS. We should clarify that it is possible that authors were using these assessments/screening criteria in the actual trial, but they may not have been reporting these measures and we can only extract and analyze the descriptive statistics based on the published information. Indeed, many inclusion/exclusion criteria

TABLE 2 | Most common data elements for baseline measures in the 489 independent groups included in SCOAR.

Measure	Number of groups with mean/median value reported	Number of groups with cut-off stated in I/E criteria only	Not reported
MMSE (or modified) (37)	82 (17%)	70 (14%)	337 (69%)
Ashworth (or modified) (38)	63 (13%)	25 (5%)	401 (82%)
Barthel (or modified) (39)	68 (14%)	2 (<1%)	419 (86%)
FAC (40)	46 (9%)	10 (2%)	433 (89%)
FIM (41)	47 (10%)	2 (<1%)	440 (90%)
Brunnstrom stages (42)	24 (5%)	20 (4%)	445 (91%)
Motricity Index (43)	28 (6%)	7 (1%)	454 (93%)
NIHSS (44)	34 (7%)	0 (0%)	455 (93%)
Berg Balance Scale (45)	34 (7%)	0 (0%)	455 (93%)

I/E, inclusion/exclusion criteria, MMSE, Mini-Mental Status Exam, FAC, Functional Ambulation Category, FIM, Functional Independence Measure, NIHSS, National Institutes of Health Stroke Scale.

were vague, with authors often reporting “no excessive spasticity” (implying perhaps the Ashworth scale was used), “no cognitive/communicative impairments” (implying the MMSE may have been used), or “no severe depression” (implying perhaps the Beck Depression Inventory or the Center for Epidemiological Studies

depression scale were used). Without more complete reporting, we have no way of knowing which measures were collected and how participants were evaluated. This presents a major gap in stroke rehabilitation knowledge (1).

Classification and Description of Therapy Types

An unexpected difficulty in the construction of the SCOAR database was in how to categorize/classify different therapeutic interventions. Ideally, researchers could use the SCOAR database to look at effect-sizes as a function of therapy-type or ask other meta-scientific questions about therapy types. Indeed, we were able to extract a short phrase or informative description of the therapy for almost all of the experimental groups (e.g., “CIMT,” “intensive progressive treadmill training”; these descriptions are listed under the “group_desc” variable in the SCOAR database), consistent with TIDieR Criteria (46, 47). For control groups, however, this proved much more difficult, with many descriptions being simply “conventional physiotherapy,” “standard care,” or “routine care.” We are currently pursuing text-mining approaches to better quantify the nature of therapy provided in control-arms of trials. In the included RCTs, however, control groups were under-described compared to experimental groups (48). Looking strictly at the Methods sections of papers, for instance, control groups only had a mean (SD) of 155 (112) words and 0.82 (1.9) references dedicated to their description. Experimental groups, conversely, had 271 (159) words and 1.77 (3.4) references dedicated to their description. Example control group descriptions are:

“All participants received standard rehabilitation, including 40 minutes each of physiotherapy and occupational therapy, given once per day, five days per week for six weeks, by the rehabilitation team. [... sentence about experimental group omitted...] The control group had visits and discussions of 20 minutes at least three times per week instead.” (49)

“All participants received the site’s conventional inpatient rehabilitation. They also performed a daily 10-m walk (or shorter distance walk until 10 m was feasible) as part of a physical therapy session.” (50)

“All participants received a duration-matched intervention for 90 to 105 minutes/day, 5 days/week for 4 weeks. The CT group received an intensive therapist-administered control therapy matched in duration with the RT groups. Occupational therapy techniques used in the treatment protocols included neurodevelopmental treatment, muscle strengthening, fine-motor training, and functional task training.” (14)

Adequate descriptions of control therapies are critical, not only because the SCOAR database spans three decades and “routine” therapy has changed considerably over that period, but also because it is impossible to compare the efficacy of a treatment across trials if the control groups to which the treatment is

compared are very different. At the moment, the under-describing of control therapies allows for very different interventions to all be categorized as “standard care” in different trials.

Classification and Common Data Elements in Outcome Measures

There is a very wide range of outcome measures in SCOAR ($n = 78$ discrete outcome names) (3). Despite this multiplicity, all of the SCOAR outcomes are measures of impairment and/or motor function. Thus, these outcomes could be analyzed together in order to get a broad view of motor recovery outcomes. We have, however, built in several default filters in SCOAR to help researchers group common outcomes more quickly. The first filter is the “outcome_extremity” variable in the database, which codes outcomes based on the involvement of the upper-extremity (*ue*) and lower-extremity (*le*). Variables that do not cleanly fit into one of these categories (e.g., the Barthel Index or other activities of daily living scales) are left blank. Within the *ue* and *le* outcomes, we further subdivide outcomes with a second-level filter. Second-level filter codes for the upper-extremity in SCOAR include the Fugl-Meyer Assessment (*fma*), Action Research Arm Test (*arat*), and Wolf Motor Function Test (*wmft*). Second-level filter codes for the lower-extremity in SCOAR include measures of *walking speed* (including the 10-m walk test and variants thereof), *walking endurance* (including the 6 min walk test and variants thereof) and *balance* (which includes the Timed Up and Go Test and the Berg Balance Scale). The number of outcomes of each type is shown in Figure 3.

Naturally, some researchers may disagree with the filters/groupings, we have created for the defaults in the SCOAR database. As such, outcomes are also coded without a filter, using the name of the outcomes from the original study (the “outcome_name” variable in the database). In many cases, this unfiltered option will be appealing because it offers the user the greatest control over what information to pool together. However, if the user does not care which subscale of the FMA is being used (e.g., the *ue* total subscale, the wrist-hand subscale, or the shoulder-elbow

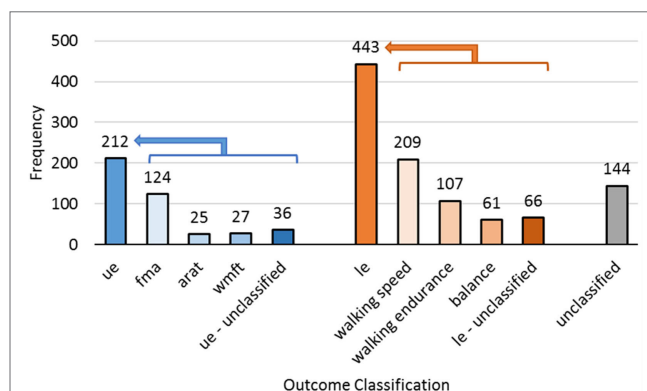


FIGURE 3 | The frequency of different outcome measures for the upper-extremity (*ue*, $n = 212$), lower-extremity (*le*, $n = 443$), and subscales of those two codes. Complex whole body assessments that did not fit into one of these two categories were unclassified ($n = 144$).

subscale) then the second-level filter *fma* would be appropriate to use. Similarly, walking speed was measured in many different ways across the different studies. For example, walking speed was measured over 10, 8, or even 4.2 m, and at both self-selected velocity and maximal effort, but if the user is not concerned about such differences in these outcomes, then the second-level filter *speed* would be appropriate to use.

The wide variation in outcomes measures means that if a researcher wants to focus on a single outcome, it would greatly reduce the amount of data that can contribute to any single analysis. As with baseline CDEs, the reporting of so many different outcome measures with so few common outcome measures imposes severe limitations on hypothesis generation and data exploration. From the SCOAR database, there is a strong argument for consistent CDEs describing participant demographics, study methodology, and clinical outcomes. Panels of experts would be required to determine exactly what these CDEs should be (e.g., What are the best UE/LE outcome measures with respect to their sensitivity and clinical application?). While the exact battery of CDEs needs to be determined, it is clear that creating a comprehensive set of rehabilitation specific CDEs would enable researchers to ask innovative questions of existing data, bringing the results of hundreds of trials to bear on clinically relevant topics with unprecedented precision and statistical power.

Data-Driven Searches and Novel Filters

A major advantage of organizing trial data in a central repository is that it can complement traditional bibliographic databases, such as PubMed or EMBASE. Traditional bibliographic tools assist authors in finding data from the top-down. That is, researchers arrive at data via well-designed searches using key terms, author names, medical subject headings, etc. By contrast, data-driven searches allow authors to find resources from the bottom-up. That is, SCOAR allows researchers to plot the relationships between variables and then click on large, outlying, or otherwise interesting data-points to obtain more information about that trial, such as patient demographics, type of intervention, and the reference for the published manuscript. For example, see <http://tinyurl.com/SCOAR-datasearch>.

In addition to these data driven searches, SCOAR enables researchers to constrain their searches by filters that are not available in bibliographic databases. For instance, a researcher could filter SCOAR to find only those trials that have used the ARAT as an outcome, in patients < 70 years old, and with sample sizes greater than 30. [Indeed, there are currently two studies in SCOAR that meet these specific criteria (51, 52).] Having these key variables describing participant demographics, research methodology, and clinical outcomes in an easily searchable database makes SCOAR a very useful compliment to existing bibliographic databases.

DISCUSSION

While SCOAR is a large step forward for data integration in stroke rehabilitation research, there are limitations associated with the database in its current form and limitations in using a database

with “group-level” statistics more generally. One limitation is completeness; the data currently in SCOAR represent two waves of data extraction: the first wave extracted the first primary or secondary outcome measure that met our inclusion criteria while the second wave extracted any assessment based on the FMA or a measure of gait speed/endurance. This creates a representative, but not exhaustive extraction of all of the outcome measures used across the various studies. Currently, we are extracting the remaining outcomes from these RCTs, regardless of whether or not this outcome was primary or secondary. Once this third wave of data extraction is complete, SCOAR will contain all of the available summary statistics for any clinical measure of impairment or function used in these studies. We have also re-implemented our systematic search to update the database through 2016 and are in the process of extracting demographic, methodological, and statistical information from those studies.

Even as the evidence base in SCOAR expands, there is a major concern with the lack of CDEs in both outcome and baseline measures. From an information architecture perspective, reporting multiple outcomes per trial poses a unique problem. For instance, self-selected gait speed and maximal gait speed or various subscales of the FMA could all be reported in the same RCT. In SCOAR, this can be resolved by creating separate rows for each outcome, which is a relatively simple problem of *scale*. For the researcher, however, it is not always clear which measure is best or if one should average across multiple related measures, reflecting a more complicated problem of *ontology* (i.e., what are the fundamental constructs in rehabilitation and how do we measure them?) (53–55). This lack of CDEs is indicative of a larger problem for rehabilitation science as a field: researchers need to think critically about which measures to collect/analyze in order to reduce the risk of false positives (56–58) and (ideally) develop a set of common measures for key constructs (1, 2, 59) that would allow for greater data integration across trials.

Similarly, a major concern across trials was in the way that therapies were reported. For control therapies in particular, the descriptions of the control interventions were vague and under-reported in comparison to experimental interventions. Control therapies had approximately half of the words dedicated to their description as experimental therapies. This lack of detail might be acceptable if control therapies had sufficient references to support them (e.g., references to standard operating procedures or other published guidelines). However, this does not seem to be the case, as control interventions had about half of the number of references in their descriptions compared to experimental interventions, and less than one reference on average. (It should be noted that these word/reference counts were based only on the text in the Methods sections of the original papers.)

For both control and experimental therapies, there is also a concern about how the dose of therapy is reported across trials. At the moment, the only consistently reported measure was time scheduled for therapy and we stress that this measure gives only a rough indication of the amount of physical practice actually performed during therapy. It would be preferable to develop a system where the type, intensity, and volume of physical practice could be tracked for individual patients and consistently



(5) design clinical trials by enabling more accurate and comprehensive power analyses.

AUTHOR CONTRIBUTIONS

KL, LB, and CL designed the systematic review and conceived the creation of the database. KL, SS, and AR oversaw the screening of manuscripts and data extraction. KL wrote the R code for management and analysis of the data. All authors contributed to the writing of the manuscript and approved the final version for publication.

REFERENCES

- Bernhardt J, Borschmann K, Boyd L, Thomas Carmichael S, Corbett D, Cramer SC, et al. Moving rehabilitation research forward: developing consensus statements for rehabilitation and recovery research. *Int J Stroke* (2016) 11(4):454–8. doi:10.1177/1747493016643851
- Saver JL, Warach S, Janis S, Odenkirchen J, Becker K, Benavente O, et al. Standardizing the structure of stroke clinical and epidemiologic research data: the National Institute of Neurological Disorders and Stroke (NINDS) Stroke Common Data Element (CDE) project. *Stroke* (2012) 43(4):967–73. doi:10.1161/STROKEAHA.111.634352
- Ali M, English C, Bernhardt J, Sunnerhagen KS, Brady M; VISTA-Rehab Collaboration. More outcomes than trials: a call for consistent data collection across stroke rehabilitation trials. *Int J Stroke* (2013) 8(1):18–24. doi:10.1111/j.1747-4949.2012.00973.x
- Wickham H. Tidy data. *J Stat Softw* (2014) 59(10):1–23. doi:10.18637/jss.v059.i10
- Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. *PLoS One* (2014) 9(2):e87987. doi:10.1371/journal.pone.0087987
- Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke* (2014) 45(7):2053–8. doi:10.1161/STROKEAHA.114.004695
- Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* (2007) 7(1):16. doi:10.1186/1472-6947-7-16
- Aisen ML, Krebs HI, Hogan N, McDowell F, Volpe BT. The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke. *Arch Neurol* (1997) 54(4):443–6. doi:10.1001/archneur.1997.00550160075019
- Volpe B, Krebs H, Hogan N, Edelstein L, Driels C. Robot training enhances motor outcome in patients with stroke maintained over 3 years. *Neurology* (1999) 53:1874–6. doi:10.1212/WNL.53.8.1874
- Dean CM, Ada L, Bampton J, Morris ME, Katrak PH, Potts S. Treadmill walking with body weight support in subacute non-ambulatory stroke improves walking capacity more than overground walking: a randomised trial. *J Physiother* (2010) 56(2):97. doi:10.1016/S1836-9553(10)70039-4
- Ada L, Dean C, Morris D, Simpson J, Katrak PH. Randomized trial of treadmill walking with body weight support to establish walking in subacute stroke: the MOBILISE trial. *Stroke* (2010) 41:1237. doi:10.1161/STROKEAHA.109.569483
- Lohse K, Boyd L, Lang C. Centralized open-access research (COAR): a database for stroke rehabilitation. *Poster Presented at the 2015 Meeting of the American Society for Neurorehabilitation*. Chicago, IL (2015).
- Volpe BT, Krebs HI, Hogan N, Edelstein OTRL, Diels C, Aisen M. A novel approach to stroke rehabilitation: robot-aided sensorimotor stimulation. (2000) 54(10):1938–44.
- Hsieh Y, Wu C, Lin K, Yao G, Wu K, Chang Y. Dose-response relationship of robot-assisted stroke motor rehabilitation: the impact of initial motor status. *Stroke* (2012) 43(10):2729–34. doi:10.1161/STROKEAHA.112.658807
- Westlake KP, Patten C. Pilot study of Lokomat versus manual-assisted treadmill training for locomotor recovery post-stroke. *J Neuroeng Rehabil* (2009) 6:18. doi:10.1186/1743-0003-6-18

ACKNOWLEDGMENTS

The authors would like to thank Anica Villamayor, Nicholas Snow, Lauren Brewer, and Taylor Buchanan for their assistance screening articles by title and abstract.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fneur.2016.00153>

- Werner C, Von Frankenberg S, Treig T, Konrad M, Hesse S. Treadmill training with partial body weight support and an electromechanical gait trainer for restoration of gait in subacute stroke patients: a randomized crossover study. *Stroke* (2002) 33(12):2895. doi:10.1161/01.STR.0000035734.61539.F6
- Lang CE, Macdonald JR, Reisman DS, Boyd L, Jacobson Kimberley T, Schindler-Ivens SM, et al. Observation of amounts of movement practice provided during stroke rehabilitation. *Arch Phys Med Rehabil* (2009) 90(10):1692–8. doi:10.1016/j.apmr.2009.04.005
- Wolf SL, Winstein CJ, Miller JP, Taub E, Uswatte G, Morris D, et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA* (2006) 296(17):2095–104. doi:10.1001/jama.296.17.2095
- Dromerick AW, Lang CE, Birkenmeier RL, Wagner JM, Miller JP, Videen TO, et al. Very early constraint-induced movement during stroke rehabilitation (VECTORS): a single-center RCT. *Neurology* (2009) 73(3):195–201. doi:10.1212/WNL.0b013e3181ab2b27
- Fuzaro AC, Guerreiro CT, Galetti FC, Jucá RB, Araujo JE. Modified constraint-induced movement therapy and modified forced-use therapy for stroke patients are both effective to promote balance and gait improvements. *Rev Bras Fisioter* (2012) 16(2):157–65. doi:10.1590/S1413-35552012005000010
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons (2011).
- Ada L, Dean CM, Lindley R. Randomized trial of treadmill training to improve walking in community-dwelling people after stroke: the AMBULATE trial. *Int J Stroke* (2013) 8(6):436–44. doi:10.1111/j.1747-4949.2012.00934.x
- Bale M, Strand LI. Does functional strength training of the leg in subacute stroke improve physical performance? A pilot randomized controlled trial. *Clin Rehabil* (2008) 22(10–11):911–21. doi:10.1177/0269215508090092
- Brunner IC, Skouen JS, Strand LI. Is modified constraint-induced movement therapy more effective than bimanual training in improving arm motor function in the subacute phase post stroke? A randomized controlled trial. *Clin Rehabil* (2012) 26(12):1078–86. doi:10.1177/0269215512443138
- Chan MK, Tong RK, Chung KY. Bilateral upper limb training with functional electric stimulation in patients with chronic stroke. *Neurorehabil Neural Repair* (2009) 23(4):357–65. doi:10.1177/1545968308326428
- Galvin R, Cusack T, O'Grady E, Murphy TB, Stokes E. Family-mediated exercise intervention (FAME): evaluation of a novel form of exercise delivery after stroke. *Stroke* (2011) 42(3):681–6. doi:10.1161/STROKEAHA.110.594689
- Janssen TW, Beltman JM, Elich P, Koppe PA, Konijnenbelt H, de Haan A, et al. Effects of electric stimulation-assisted cycling training in people with chronic stroke. *Arch Phys Med Rehabil* (2008) 89(3):463–9. doi:10.1016/j.apmr.2007.09.028
- Jung J. Effects of virtual-reality treadmill training on balance and balance self efficacy in stroke patients with a history of falling. *J Phys Ther Sci* (2012) 24(5):1133–6. doi:10.1589/jpts.24.1133
- Kuys SS, Brauer SG, Ada L. Higher-intensity treadmill walking during rehabilitation after stroke in feasible and not detrimental to walking pattern or quality: a pilot randomized trial. *Clin Rehabil* (2011) 25(4):316–26. doi:10.1177/0269215510382928
- Sullivan KJ, Brown DA, Klassen T, Mulroy S, Ge T, Azen SP, et al. Effects of task-specific locomotor and strength training in adults who were ambulatory after stroke: results of the STEPS randomized clinical trial. *Phys Ther* (2007) 87(12):1580–607. doi:10.2522/ptj.20060310

31. In TS, Jung KS, Lee SW, Song CH. Virtual reality reflection therapy improves motor recovery and motor function in the upper extremities of people with chronic stroke. *J Phys Ther Sci* (2012) 24(4):339–43. doi:10.1589/jpts.24.339
32. Wang RY, Chen HI, Chen CY, Yang YR. Efficacy of Bobath versus orthopaedic approach on impairment and function at different motor recovery stages after stroke: a randomized controlled study. *Clin Rehabil* (2005) 19(2):155–64. doi:10.1191/0269215505cr850oa
33. Yang Y, Yen J, Wang R-Y, Yen L, Lieu F-K. Gait outcomes after additional backward walking training in patients with stroke: a randomized controlled trial. *Clin Rehabil* (2005) 19(3):264–73. doi:10.1191/0269215505cr860oa
34. Yang Y, Wang R, Lin K, Chu M, Chan R. Task-oriented progressive resistance strength training improves muscle strength and functional performance in individuals with stroke. *Clin Rehabil* (2006) 20:860–70. doi:10.1177/0269215506070701
35. Duncan PW, Sullivan KJ, Behrman AL, Azen SP, Wu SS, Nadeau SE, et al. Body-weight-supported treadmill rehabilitation after stroke. *N Engl J Med* (2011) 364(21):2026. doi:10.1056/NEJMoa1010790
36. Stinear CM, Barber PA, Coxon JP, Fleming MK, Byblow WD. Priming the motor system enhances the effects of upper limb therapy in chronic stroke. *Brain* (2008) 131:1381–90. doi:10.1093/brain/awn051
37. Teng EL, Chui HC. The modified mini-mental state (3MS) examination. *J Clin Psychiatry* (1987) 48(8):314–8.
38. Bohannon RW, Smith MB. Interrater reliability of a modified ashworth scale of muscle spasticity. *Phys Ther* (1987) 67(2):206–7.
39. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* (1989) 42(8):703–9. doi:10.1016/0895-4356(89)90065-6
40. Williams G. Functional ambulation classification. *Encycl Clin Neuropsychol* (2011) 1105–6. doi:10.1007/978-0-387-79948-3_1935
41. Ottenbacher K, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. *Arch Phys Med Rehabil* (1996) 77(12):1226–32. doi:10.1016/S0003-9993(96)90184-7
42. Brunnstrom S. Motor testing procedures in hemiplegia: based on sequential recovery stages. *Phys Ther* (1966) 46(4):357–75.
43. Cameron D, Bonhannon RW. Criterion validity of lower extremity Motricity Index scores. *Clin Rehabil* (2000) 14(2):208–11. doi:10.1191/026921500675786655
44. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol* (1989) 46(6):660–2. doi:10.1001/archneur.1989.00520420080026
45. Berg K, Wood-Dauphinee S, Williams JL. The balance scale: reliability assessment with elderly residents and patients with an acute stroke. *Scand J Rehabil Med* (1995) 27(1):27–36.
46. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* (2014) 348(7659):g1687. doi:10.1136/bmj.g1687
47. Yamato T, Maher C, Saragiotto B, Moseley A, Hoffmann T, Elkins M, et al. The TIDieR checklist will benefit the physical therapy profession. *Phys Ther* (2016) 96(7):930–1. doi:10.2522/ptj.2016.96.7.930
48. Hoffmann TC, Walker MF. “TIDieR-ing up” the reporting of interventions in stroke research: the importance of knowing what is in the “black box. *Int J Stroke* (2015) 10(5):657–8. doi:10.1111/ijls.12524
49. Chen J-C, Lin C-H, Wei Y-C, Hsiao J, Liang C-C. Facilitation of motor and balance recovery by thermal intervention for the paretic lower limb of acute stroke: a single-blind randomized clinical trial. *Clin Rehabil* (2011) 25(9):823–32. doi:10.1177/0269215511399591
50. Dobkin BH, Plummer-D’Amato P, Elashoff R, Lee J, Sirrows G. International randomized clinical trial, stroke inpatient rehabilitation with reinforcement of walking speed (SIRROWS), improves outcomes. *Neurorehabil Neural Repair* (2010) 24(3):235. doi:10.1177/1545968309357558
51. Morris JH, Van Wijck F, Joice S, Ogston SA, Cole I, MacWalter RS. A comparison of bilateral and unilateral upper-limb task training in early poststroke rehabilitation: a randomized controlled trial. *Arch Phys Med Rehabil* (2008) 89(7):1237–45. doi:10.1016/j.apmr.2007.11.039
52. van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. *Stroke* (1999) 30(11):2369–75. doi:10.1161/01.STR.30.11.2369
53. Levin M, Kleim J, Wolf S. What do motor “recovery” and “compensation” mean in patients following stroke? *Neurorehabil Neural Repair* (2009) 23:313–9. doi:10.1177/1545968308328727
54. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair* (2008) 22(1):78–90. doi:10.1177/1545968307305353
55. See J, Dodakian L, Chou C, Chan V, McKenzie A, Reinkensmeyer DJ, et al. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. *Neurorehabil Neural Repair* (2013) 27(8):732–41. doi:10.1177/1545968313491000
56. Ioannidis J. Why most published research findings are false. *PLoS Med* (2005) 2(8):e124. doi:10.1371/journal.pmed.0020124
57. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* (2002) 2(1):1. doi:10.1186/1471-2288-2-8
58. Lohse KR, Buchanan TL, Miller MW. Under-powered and over-worked: problems with data analysis in motor learning studies. *J Mot Learn Dev* (2016) 4(1):37–58. doi:10.1123/jmld.2015-0010
59. Ali M, Fulton R, Quinn T, Brady M; VISTA Collaboration. How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. *Stroke* (2013) 44(11):3161–5. doi:10.1161/STROKEAHA.113.001126
60. Lang C, Lohse K, Birkenmeier RL. Dose and timing in neurorehabilitation: prescribing motor therapy after stroke. *Curr Opin Neurol* (2015) 28:549–55. doi:10.1097/WCO.0000000000000256

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Lohse, Schaefer, Raikes, Boyd and Lang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.